CENTRE FOR
SYSTEMS
SCIENCE

LCCR

Simon Fraser University, Burnaby, B.C., Canada V5A 1S6 / (604) 291-4704

Simple Estimation of the Convergence Rate of a Markov Chain*

Slawomir Pilarski and Tiko Kameda

CSS/LCCR TR90-14

June 1990

92-05797/

Simple Estimation of the Convergence Rate of a Markov Chain*

Slawomir Pilarski† and Tiko Kameda†

Abstract

We present a simple method for estimating an upper bound on the convergence rate of an ergodic markov chain, which often produces a much tighter bound than those proposed previously. The complexity of computing the bound is linear in the input size, if the input is given in terms of a transition matrix.

*This work was partially supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Advanced Systems Institute of British Columbia.

†School of Computing Science, Simon Fraser University, Burnaby, B.C., Canada V5A 1S6

1 Introduction

Markov processes are useful to model wide classes of systems and recently there has been considerable interest in Computer Science in analyzing the convergence rate of ergodic Markov chains[1] [4, 8, 5, 6, 9].

The traditional approach to the analysis of convergence rate is based on the eigenvalues of the transition matrix (see e.g., [10, 3]). Some limitations of this approach are discussed in [9]. Several recent papers analyze the rate of convergence using some other concepts such as *coupling* [1] and *stopping times* [2]. They have the disadvantage that the analysis gets very complicated for processes that lack high degree of symmetry. That was an incentive to find simpler upper bounds on the convergence rate by analyzing the transition graph [8] [9].

Let us consider an n-state ergodic Markov process whose transition matrix is given by

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & \cdots & p_{1,n} \\ \vdots & \vdots & \vdots \\ p_{n,1} & \cdots & p_{n,n} \end{bmatrix},$$

where $p_{i,j}$ is the probability of transition from state i to state j. Let $\vec{x}(t) = [x_1(t), \ldots, x_n(t)]$ denote a vector of probabilities such that $x_i(t)$ is the probability of being in state i at time t. Thus, for $t = 0, 1, \ldots$, we have $\sum_{i=1}^{n} x_i(t) = 1$ and $\vec{x}(t+1) = \vec{x}(t)\mathbf{P}$. The stationary distribution is denoted by $\vec{\pi} = [\pi_1, \ldots, \pi_n]$.

[1] There does not seem to be a standard definition of Markov chain and Markov process in the literature. We use them more or less interchangeably.

To discuss the convergence rate, we must first introduce a measure of "distance" (or "discrepancy") between the distribution of state probabilities at time t and the stationary distribution. Sinclair and Jerrum [11], [8] adopt as such a measure the largest relative difference between $\vec{x}(t)$ and π, i.e.,

$$\max_{1 \le i \le n} \frac{|x_i(t) - \pi_i|}{\pi_i}.$$

This measure, maximized over all possible initial distributions, is denoted $\Delta(t)$ and can be expressed as

$$\Delta(t) = \max_{1 \le i \le n, 1 \le j \le n} \frac{|p_{i,j}(t) - \pi_i|}{\pi_i}, \tag{1}$$

where $p_{i,j}(t)$ is the (i,j) entry of $\mathbf{P}^t$. Mihail [9], on the other hand, adopts two different measures: the square of the euclidian distance,

$$\sum_{i=1}^{n} (x_i(t) - \pi_i)^2,$$

and the following sum:

$$\sum_{i=1}^{n} \frac{(x_i(t) - \pi_i)^2}{\pi_i},$$

that in [9] is denoted by $\|e(t)\|$.

Simple upper bounds are established in [8] and [9] on the convergence rate, using variants of the notion of "conductance" first defined in [11] for **time-reversible** ergodic Markov chains, i.e., those satisfying $\pi_i p_{i,j} = \pi_j p_{j,i}$ for any i and j. To be more specific, the **conductance**, denoted Φ, is defined as follows:

$$\Phi = \min_{S \mid \sum_{i \in S} \pi_i \le 1/2} \frac{\sum_{i \in S} \sum_{j \notin S} \pi_i p_{i,j}}{\sum_{i \in S} \pi_i},$$

where S is a subset of states. The main result of [8] concerning convergence rate states that for any time-reversible and **strongly aperiodic**[2] ergodic Markov chain

$$(\min_{1 \le i \le n} \pi_i)\Delta(t) \le (1 - \Phi^2/2)^t$$

holds. Mihail [9] has derived a similar bound on the convergence rate of $\|e(t)\|$ for time-reversible, strongly aperiodic ergodic Markov chains:

$$\|e(t)\| \le (1 - \Phi^2)^t \, \|e(0)\|.$$

It should be noted, however, that the Jerrum-Sinclair bound is related to distance while Mihail's improved bound is related to the square of distance. It follows from the Jerrum-Sinclair bound that $(1 - \Phi^2/2)^2 = 1 - \Phi^2 + \Phi^4/4$ constitutes an upper bound on the convergence rate for the measure used in [9]. Thus, the improvement over [8] is not as significant as it might appear.

Another bound presented in [9] applies in general to any ergodic Markov chain and gives

$$\|e(t)\| \le (1 - \Phi_*^2/2)^t \, \|e(0)\|,$$

where Φ_* is the **merging conductance** defined as follows:

$$\Phi_* = \min_{S | \sum_{i \in S} \pi_i \le 1/2} \frac{\sum_{j \in S} \sum_{k \notin S} \sum_{i=1}^{n} \frac{\pi_j p_{j,i} \pi_k p_{k,i}}{\pi_i}}{\sum_{i \in S} \pi_i}$$

Unfortunately, determining the conductance and the merging conductance may be computationally expensive. (There is no analysis of computational complexity in

[2]This means $\min_{1 \le i \le n} p_{i,i} \ge 1/2$.

[8] or [9].) Moreover, the upper bounds they provide are often far from tight, as seen from the example presented in section 3. Therefore, there is a lot of room for improvement in both of these directions. In the next section, we present our new upper bound. In section 3 we will compare the performance of our method with that of the others in terms of an example.

2 A New Method of Upper Bound Estimation

As a measure of distance between $\vec{x}(t)$ and $\vec{\pi}$ we use

$$\delta(\vec{x}(t)) = \max_{1 \le i \le n} |x_i(t) - \pi_i|. \tag{2}$$

It is easy to see that the following simple relation exists between $\delta(\vec{x}(t))$ and $\Delta(t)$ defined in eq.(1):

$$(\min_i \pi_i)\Delta(t) \le \max_{\vec{x}(0)} \delta(\vec{x}(t)), \tag{3}$$

where max is over all initial probability distributions $\vec{x}(0)$.

Theorem *Given the transition matrix* $\mathbf{P}$ *of any ergodic Markov process, let* $\{q_1^{[k]}, \ldots, q_n^{[k]}\}$ *be a sorted list of the elements in the* k*-th column of* $\mathbf{P}$*, such that for any* $i, 1 \le i \le n-1$, $q_i^{[k]} \ge q_{i+1}^{[k]}$. *For any probability distribution vector,* $\vec{x}(t)$,

$$\delta(\vec{x}(t+1)) \le \rho(\mathbf{P})\delta(\vec{x}(t)),$$

where

$$\rho(\mathbf{P}) = \max_{1 \le k \le n} \Big(\sum_{i=1}^{\lfloor n/2 \rfloor} q_i^{[k]} - \sum_{i=\lceil n/2 \rceil + 1}^{n} q_i^{[k]} \Big).$$

Proof: Since $\vec{x}(t+1) = \vec{x}(t)\mathbf{P}$, by (2), there exists an integer m such that $1 \leq m \leq n$ and

$$\delta(\vec{x}(t+1)) = |\sum_{i=1}^{n} x_i(t)p_{i,m} - \pi_m|.$$

Since $\vec{\pi} = \vec{\pi}\mathbf{P}$, substituting $\pi_m = \sum_{i=1}^{n} \pi_i p_{i,m}$ into the above equation, we obtain

$$\delta(\vec{x}(t+1)) = |\sum_{i=1}^{n} p_{i,m}(x_i(t) - \pi_i)|.$$

Because $\sum_{i=1}^{n}(x_i(t) - \pi_i) = 0$, this equation can be transformed into

$$\delta(\vec{x}(t+1)) = |\sum_{i=1}^{n}(p_{i,m} - c)(x_i(t) - \pi_i)|,$$

where c is any constant. Thus, inequality

$$\delta(\vec{x}(t+1)) \leq \sum_{i=1}^{n} |p_{i,m} - c||x_i(t) - \pi_i|,$$

holds for any c. From definition of $\delta(\vec{x})$, it follows that

$$\delta(\vec{x}(t+1)) \leq \sum_{i=1}^{n} |p_{i,m} - c|\delta(\vec{x}(t)). \tag{4}$$

By definition of $\{q_i^{[m]}\}$, we have

$$\min_c \sum_{i=1}^{n} |p_{i,m} - c| = \min_c \sum_{i=1}^{n} |q_i^{[m]} - c|.$$

It's a simple fact that expression $\sum_{i=1}^{n} |q_i^{[m]} - c|$ takes the minimum value for c such that

$$\begin{cases} c = q^{[m]}_{\lceil n/2 \rceil} & \text{if } n \text{ is odd} \\ q^{[m]}_{(n/2)} < c < q^{[m]}_{(n/2+1)} & \text{if } n \text{ is even} \end{cases}$$

This implies that

$$\min_c \sum_{i=1}^{n} |q_i^{[m]} - c| = \sum_{i=1}^{\lfloor n/2 \rfloor} q_i^{[m]} - \sum_{i=\lceil n/2 \rceil + 1}^{n} q_i^{[m]},$$

and inequality (4) can be transformed into

$$\delta(\vec{x}(t+1)) \leq (\sum_{i=1}^{\lfloor n/2 \rfloor} q_i^{[m]} - \sum_{i=\lceil n/2 \rceil + 1}^{n} q_i^{[m]})\delta(\vec{x}(t)).$$

□

From the above theorem, it immediately follows that, for $t = 0, 1, \ldots,$

$$\delta(\vec{x}(t)) \leq \beta(t)\delta(\vec{x}(0)), \tag{5}$$

where $\beta(t) = \rho^t(\mathbf{P})$.

Before discussing the limitations of our approach and how to overcome some of them, we now present an example.

3 An Example

Consider the following symmetric transition matrix, where $0 < p \leq 1/2$.

$$\mathbf{P} = \begin{bmatrix} 1-2p & p & 0 & 0 & p \\ p & 1-2p & p & 0 & 0 \\ 0 & p & 1-2p & p & 0 \\ 0 & 0 & p & 1-2p & p \\ p & 0 & 0 & p & 1-2p \end{bmatrix}$$

The characteristic polynomial of this matrix is

$$(1-\lambda)(1-\frac{5-\sqrt{5}}{2}p-\lambda)^2(1-\frac{5+\sqrt{5}}{2}p-\lambda)^2.$$

Thus, the rate of convergence that is implied by the eigenvalues equals

$$\begin{cases} 1 - \frac{5-\sqrt{5}}{2}p & \text{for } 0 < p \leq 2/5 \\ \frac{5+\sqrt{5}}{2}p - 1 & \text{for } 2/5 < p \leq 1/2. \end{cases} \tag{6}$$

For this $\mathbf{P}$, we have

$$\rho(\mathbf{P}) = \begin{cases} (1-2p+p) - 0 = 1 - p & \text{for } 0 < p \leq 1/3 \\ (p+p) - 0 = 2p & \text{for } 1/3 < p \leq 1/2. \end{cases}$$

It is easy to find that the equilibrium probabilities

$$\pi_i = 1/5 \text{ for all } i.$$

Using relation (3) and formula (5), we can obtain an upper bound on $(\min_i \pi_i)\Delta(t)$.

$$(\min_i \pi_i)\Delta(t) \leq \max_{\vec{x}(0)} \beta(t)\delta(\vec{x}(0)) = \begin{cases} \frac{4}{5}(1-p)^t & \text{for } 0 < p \leq 1/3 \\ \frac{4}{5}(2p)^t & \text{for } 1/3 < p \leq 1/2. \end{cases} \tag{7}$$

Since, for $0 < p \leq 1/4$, $\mathbf{P}$ represents a time-reversible, strongly aperiodic ergodic Markov process, one can compute the Jerrum-Sinclair upper bound. Note that conductance Φ is defined by the set $S = \{1, 2\}$ and $\Phi = p$, which gives us the following upper bound:

$$(\min_i \pi_i)\Delta(t) \leq (1 - \Phi^2/2)^t = (1 - p^2/2)^t \text{ for } 0 < p \leq 1/4. \tag{8}$$

As stated in the Introduction, a result in [9] now gives us

$$\|e(t)\| \leq (1-p^2)^t \|e(0)\| \text{ for } 0 < p \leq 1/4. \tag{9}$$

Another result in [9], using the merging conductance, yields

$$\|e(t)\| \leq \begin{cases} (1 - 2p^2(1-p)^2)^t \|e(0)\| & \text{for } 0 < p \leq 2/5 \\ (1 - p^2(4-7p)^2/2)^t \|e(0)\| & \text{for } 2/5 < p \leq 1/2. \end{cases} \tag{10}$$

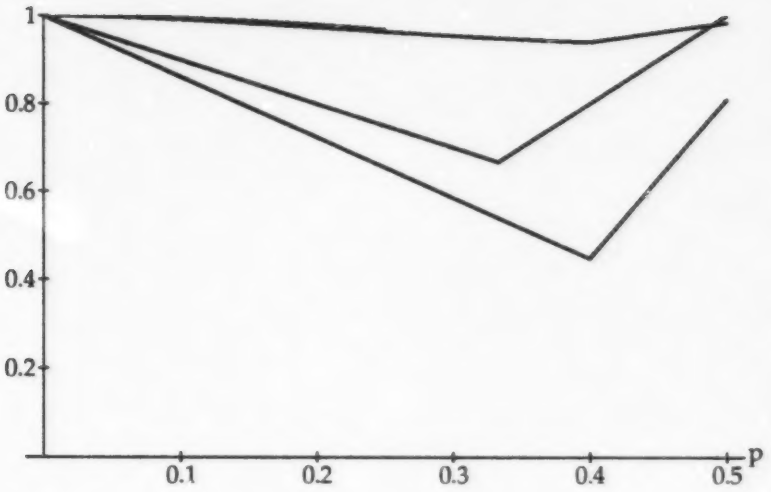

Figure 1: Upper bounds on convergence rate as functions of p.

In Figure 1, we have plotted the five rates given in formulae (6),(7), (8), (9) and (10) for $0 < p \leq 0.5$. The bottom curve shows the eigenvalues given in (6), while the second from the bottom shows the bound on the rate given in (7). The Jerrum-Sinclair bound is plotted for $0 < p \leq 0.25$ and is just slightly above the bounds of Mihail, (9) and (10). To perform a fair comparison we have used the square roots of the rates given in (9) and (10). It is seen from the figure that for the values of p in the range $0 < p \leq 0.49$, (7) gives the best bound. However, for a small range of p, i.e., $0.49 < p \leq 0.5$, (10) gives a better bound. Note also that convergence rates given by (8) and (9) are practically indistinguishable, and only slightly differ from the rate given by (10).

4 Conclusions and Discussions

Our example in the previous section was selected to satisfy all of the following criteria:

1. It should be parametrized (i.e., by p) so that it represents a family of examples instead of just one numerical example.

2. The transition matrix $\mathbf{P}$ should be diagonalizable, so the second largest eigenvalue of $\mathbf{P}$ bounds the convergence rate.

3. For certain values of p, $\mathbf{P}$ should be strongly aperiodic and time-reversible, so that Jerrum and Sinclair's bound can be evaluated.

4. For certain values of p, $\mathbf{P}$ should *not* be strongly aperiodic, so that Mihail's bound can be compared with the new bound for such cases.

It is emphasized that we do not imply that the above example is "typical" in any sense.

From the above example and others that we have tested, it is seen that our estimation method, although very simple, can be at the same time very effective.

We now analyze the computational complexity of the method presented in this paper. Our method requires that the median of each column of the given transition matrix $\mathbf{P}$ be found. This can be done in $O(n)$ time for each column [7], hence in $O(n^2)$ time for all the n columns. Clearly, this time dominates the additional computation

we need to obtain our upper bound. A sparse matrix is often represented by an adjacency structure consisting of a set of linked lists, one for each column. The linked list for a column contains all and only the non-zero elements in the column. If the input is given in terms of such an adjacency structure, then instead of the median we need to compute the kth largest element in each column, where $k = \lceil n/2 \rceil$. This can be done in linear time in the size of the linked list for the column [7]. Thus the total time is still linear in the input size.

The time complexity of computing the conductance and merging conductance has not been analyzed. However, it appears that they are expensive to compute.

In order for the Theorem to be useful, we must have $\rho(\mathbf{P}) < 1$. Given a transition matrix $\mathbf{P}$ of an ergodic Markov process, if $\mathbf{P}$ doesn't satisfy this condition, then one can compute $\mathbf{P}^2$, $\mathbf{P}^3$, ..., until $\rho(\mathbf{P}^k) < 1$ holds for some k. Now the Theorem is applicable to $\mathbf{P}^k$, and $\rho(\mathbf{P}^k)$ is an upper bound on the convergence rate for every k steps.

It is not difficult to show that for any matrix $\mathbf{P}$ with $\rho(\mathbf{P}) \geq 1$, there exists a vector $\vec{x} \neq \vec{\pi}$ such that $\delta(\vec{x}) \leq \delta(\vec{x}\mathbf{P})$. This means that $\rho(\mathbf{P}) < 1$ is a necessary and sufficient condition that $\delta(\vec{x}(t))$ shrinks after each transition regardless of $\vec{x}(t)$.

In summary, we draw the following conclusions.

- There are cases where our new bound is much better than the others.

- The amount of computation required to compute the new bound is much less than those required for the others.

- The new bound is not always the best.

References

[1] D. Aldous, Random walks on ginite groups and rapidly mixing Markov chains, *Seminaire de Probabilites XVII*, Lecture Notes in Mathematics, Vol. 986, Springer-Verlag, 1983, 243-297.

[2] D. Aldous and P. Diaconis, Shuffling cards and stopping times, *American Mathematical Monthly, Vol. 93*, 1986, 333-348.

[3] N. Alon, Eigenvalues and expanders, *Combinatorica Vol. 6(2)*, 1986, 83-96.

[4] A.Z. Broder, How hard is it to marry at random? (On the approximation of the permanent), *Proc. 18th ACM Symp. Theory of Computing (STOC)*, 1986, 50-58.

[5] P. Dagum, M. Luby, M. Mihail, and U. Vazirai, Polytopes, permanents, and graphs with large factors, *Proc. 29th IEEE Symp. Foundations of Computer Science (FOCS)*, 1988, 412-421.

[6] M. Dyer, A. Frieze, and R. Kanna, A random polynomial time algorithm for estimating volumes of convex bodies, *Proc. 20th STOC*, 1988, 375-381.

[7] M. Blum, R. Floyd, V. Pratt, R. Rivest and R. Tarjan, Time bounds for selection, *J. of Computer and System Sciences, Vol. 7*, 1972, 448-461.

[8] M.R. Jerrum and A. Sinclair, Conductance and the rapid mising property for Markov chains: the approximation of the permanent resolved, *Proc. 21st STOC*, 1989, 235-243.

[9] M. Mihail, Conductance and convergence of Markov chains - A combinatorial treatment of expanders -, *Proc. 30th FOCS*, 1989, 526-531.

[10] E. Seneta, *Non-Negative Matrices; An Introduction to Theory and Applications*, John-Wiley, 1973.

[11] A. Sinclair and M.R. Jerrum, Approximate counting, uniform generation and rapidly mixing Markov chains, To appear in *Information and Computation.*

Centre for Systems Science

• Simon Fraser University established the **Centre for Systems Science** as a provincial centre for advanced study. The objective of CSS is twofold: coordinate and combine advanced systems research at SFU, and encourage application of research results within industry in order to further the technological development of British Columbia.

• CSS has assembled three research groups to streamline its efforts. The orientation of these groups is as follows: intelligent systems, microelectronics, and computer and communications systems. Because of shared research interests, this latter group merges with LCCR (**Laboratory for Computer and Communications Research**) for the publication of technical reports.

• CSS's three groups involve over 80 skilled faculty researchers principally from SFU's Schools of Engineering Science and Computing Science, but also draw substantially on Mathematics, Physics, Geography, Philosophy, Linguistics, Psychology, Communication, and Kinesiology.